In a casino in Liverpool, there are two slot machines: if you play on the first one (machine A), you win 10% of the time. If you play on the second one (machine B), you win 20% of the time. You do not know which machine is which.

Your strategy: you assume the machines have an equal chance of being the better one, you select one of the two at random and put a coin in it. You lose the first bet. What is the probability that you selected machine B?

- a. 0.4701
- b. 0.4721
- c. 0.4715
- d. 0.4706 ✓

Consider the following dataset with features X, Y and Z

X	Y	Z
Jack	0	A
Jack	1	B
Amy	0	A
Amy	1	B
Sam	0	B
Sam	1	B

Let H denote the entropy function. Which of the following is true?

- a. H(X) = 1.585, H(Y) = 1, H(Z)=0.9183
- b. H(X) = 1.213, H(Y) = 1, H(Z)=0.1544
- c. H(X) = 1.785, H(Y) = 1, H(Z)=0.656
- d. H(X) = 1.585, H(Y) = 0, H(Z)=0.9183

In logistic regression, we assume a...

(logit = logarithm of the odds)

- a. a nonlinear relationship between continuous features and the logit of the outcome variable
- b. linear relationships between the logit of continuous features and the logit of the outcome variable
- c. linear relationships between the logit of continuous features and the outcome variable
- d. linear relationship between continuous features and the logit of the outcome variable

Let X denote the outcome of tossing a special 6 sided dice - this dice has 6 faces like a normal dice, but it is twice as likely to land on an even face as it is to land on an odd face. Compute the following probability: $P(X = 3|X > 2)$

- a. 0.2222
- b. 0.3333
- c. 0.1667
- d. 0.1121

Your dataset consists of documents, each of which may be represented as a 3 dimensional feature vector. You decide to fit a logistic regression to the data, and derive the following estimates for your weight vector: $\beta = (-\ln(2), \ln(5), -\ln(7))$. You then receive a new document $x_* = (1, -1, -1)$. Compute $P(y_* = 0 | x_*)$.

- a. 0.3343
- b. 0.3453
- c. 0.6814
- d. 0.5882

Which of the following statements about Naive Bayes is incorrect?

1. Features are equally important
2. Features are statistically dependent of one another given the class value
3. Features are statistically independent of one another given the class value
4. Features can be nominal or numeric

- a. Statement 4 is incorrect
- b. Statement 1 is incorrect
- c. Statement 3 is incorrect
- d. Statement 2 is incorrect

There are two jars (jar A and jar B).

Jar A is composed of 50% red balls, and 50% blue balls. Jar B is composed of 60% red balls, and 40% blue balls.

You play the following game: you toss a biased coin (which has probability of heads 0.8), and if it comes up heads you pick a ball randomly from jar A, otherwise you pick a ball randomly from jar B.

You play the game and end up with a blue ball, what is the probability that this blue ball came from jar A?

- a. 0.8333 ✔
- b. 0.6011
- c. 0.2355
- d. 0.5556

Your dataset consists of documents, each of which may be represented as a 3 dimensional feature vector. You decide to fit a logistic regression to the data, and derive the following estimates for your weight vector: $\beta = (-\ln(2), \ln(5), -\ln(7))$. You then receive a new document $x_* = (1, -1, -1)$. Compute $P(y_* = 0 | x_*)$.

- a. 0.3343
- b. 0.5882
- c. 0.6814
- d. 0.3453

Consider the following dataset with features X, Y and Z

X	Y	Z
0	0	0
0	1	1
1	0	0
1	1	1
2	0	1
2	1	1

Compute the entropy of $\frac{XY}{Z+1}$

- a. Undefined
- b. 1
- c. 1.2516
- d. 2.11328